

Abstract

We research the role that linguistics plays in improving the interpretation of video corpora. Our work is mainly focused both on speech transcriptions automatically taken from video recordings (ASR) and on optical character recognition outputs (OCR). Their analyses should help us to structure multimedia documents and to interpret their contents.

First of all, we report on the problems we faced while working with video corpora. They are divided into four classes: ASR/OCR inaccuracies, the annotation standards, characteristics of spontaneous spoken language, and merging of ASR and OCR data. Some of these problems are clearly interrelated so that global solutions can be proposed.

Linguistic analysis of multimodal corpora must be as simple as possible because of the mentioned problems. The corpus was annotated with a TnT POS tagger in order to extract nouns and verbs. Afterwards the documents were processed by the Schug chunker for the detection of the most basic linguistic fragments since the corpus had been delivered without punctuation.

Regarding noun phrases, named entities have proved to be the easiest way of relating a transcript to its video and of interpreting their contents. Besides, NE recognition can be easily improved by extending gazetteers. The spoken language redundancy has been also used to solve ASR inaccuracies considering repetition of words as evidence of their right transcription. A more complex annotation based on event structures is needed for the verb phrases analysis. The SESCO typology, made up of only three main event types and five argument classes, was chosen because of its simplicity and compositionality.

Our experiments show that it is desirable to have linguistic analysis when processing video corpora. Methods based on complex syntactic annotations cannot be taken into account because both the POS taggers and the ASR outputs are still far from being reliable. However, shallow semantic analysis provides us with a valuable basis for interpreting the videos.

¹ Research supported by the European Commission, contract FP6-027026, Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - K-Space, and partially funded by a MEC/Fulbright grant of the Spanish Government.

² DFKI GmbH

e-mail: manuel.alcantara@dfki.de, declerck@dfki.de